

Data Analysis

Name: _____

Period _____

Date _____

Worksheet 4.7

Circle the attribute you have been assigned:

Birth Month

Height

1. Create a boxplot for the Boys data and the Girls data at your station.

Boys

Girls

Create a dot plot:

Boys:

Girls:

Now, using the dotplot and/or the boxplot, answer the following questions:

2. Does it appear that boys or girls have the greater mean? Explain your choice.

3. Do boys or girls have the greater median?

How does this appear on the dotplot?

How does this appear on the boxplot?

4. Does it appear that boys or girls have the larger range? Explain your choice.
5. Does it appear that boys or girls have the greater MAD? Explain your choice.
6. Are there any data points that appear to be outliers? If so, which one(s)?
Use the $1.5 \cdot \text{IQR}$ rule to determine if there really are any outliers.
7. What, if anything, would happen to each of the following statistics if the largest number in your data set were removed?
- mean:
- median:
- range:
- IQR:
- MAD:

Data Analysis

Comparing Statistics

Worksheet 4.8

Name _____

Period _____ Date _____

The data sets below give the number of home runs by each player on the Bears and Wildcats during a season of the Oakmont Baseball League. Find the requested statistics for each team.

Bears:

2 3 5 7 8 12 14 18 19 21 25 28

Mean	Range	Median	Lower Quartile	Upper Quartile	IQR	MAD

~~~~~

Wildcats:

4    5    6    8    9    11    12    15    16    18    19    20

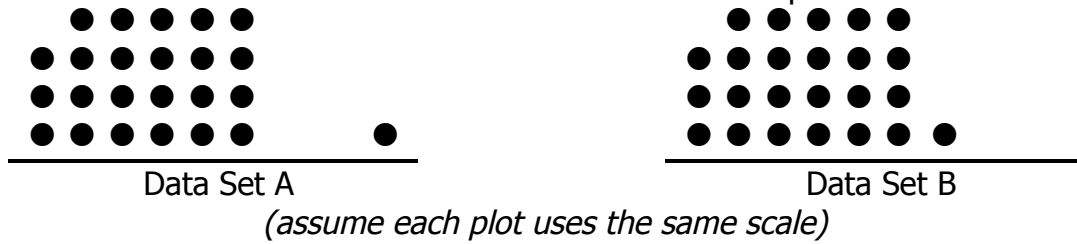
| Mean | Range | Median | Lower Quartile | Upper Quartile | IQR | MAD |
|------|-------|--------|----------------|----------------|-----|-----|
|      |       |        |                |                |     |     |

~~~~~

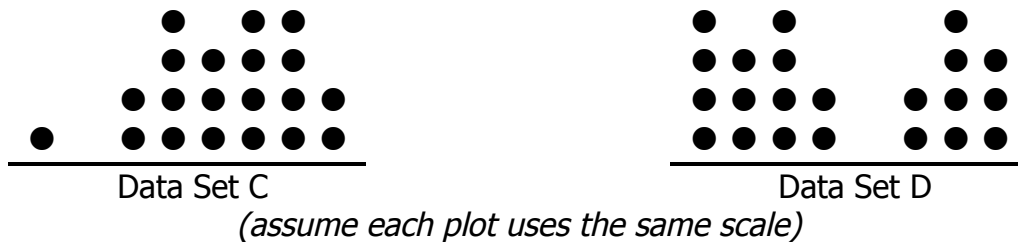
Which team is more *consistent* (all the players are similar)? How do you know?
Refer to the statistics you calculated to justify your answer.

Which team's *typical* (normal) player is better? How do you know?
Refer to the statistics you calculated to justify your answer.

Two data sets are shown below as dot plots



1. Which data set appears to have a larger mean (or do they appear to be the same)? Why?
2. Which data set appears to have a larger median (or do they appear to be the same)? Why?
3. Which data set appears to have a larger range (or do they appear to be the same)? Why?



4. Which data set appears to have a larger interquartile range (or do they appear to be the same)? Why?
5. Which data set appears to have a larger mean absolute deviation (or do they appear to be the same)? Why?

Data Analysis

Comparing Statistics

Name KEY

Period _____ Date _____

The data sets below give the number of home runs by each player on the Bears and Wildcats during a season of the Oakmont Baseball League. Find the requested statistics for each team.

Bears:
2 3 5 7 8 12 14 18 19 21 25 28

Mean	Range	Median	Lower Quartile	Upper Quartile	IQR	MAD
13.5	26	13	6	20	14	$\frac{22}{3}$ ≈ 7.333

~~~~~

Wildcats:  
4    5    6    7    8    11    12    15    16    18    19    20

| Mean | Range | Median | Lower Quartile | Upper Quartile | IQR  | MAD                                |
|------|-------|--------|----------------|----------------|------|------------------------------------|
| 12   | 16    | 12     | 6.5            | 17             | 10.5 | $\frac{62}{13}$<br>$\approx 4.769$ |

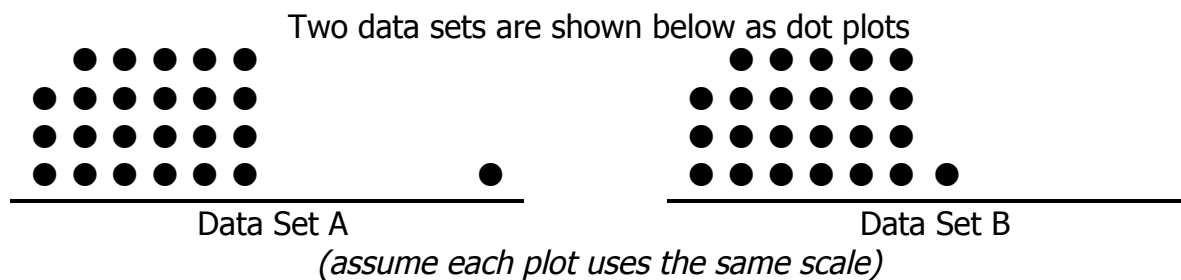
~~~~~

Which team is more *consistent* (all the players are similar)? How do you know?
Refer to the statistics you calculated to justify your answer.

Wildcats. Their range, IQR, and MAD are all lower.

Which team's *typical* (normal) player is better? How do you know?
Refer to the statistics you calculated to justify your answer.

Bears. Their mean and median are higher. Also, their upper quartile is higher, and their lower quartile is only slightly lower.



1. Which data set appears to have a larger mean (or do they appear to be the same)? Why?

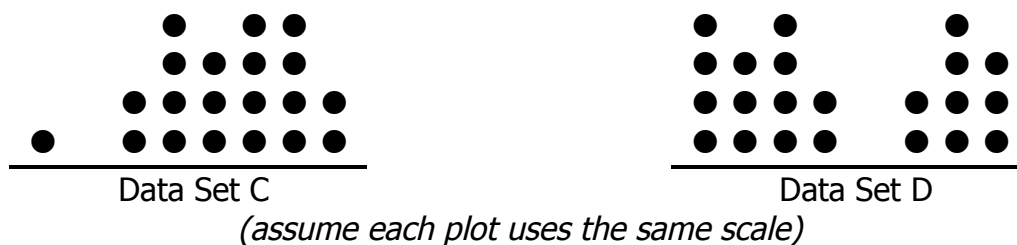
Data Set A. The outlier on the right will pull the mean up.

2. Which data set appears to have a larger median (or do they appear to be the same)? Why?

Same. The only difference is the maximum value, which does not affect the median.

3. Which data set appears to have a larger range (or do they appear to be the same)? Why?

Data Set A. The difference between the max and min is larger (wider distribution).



4. Which data set appears to have a larger interquartile range (or do they appear to be the same)? Why?

Data Set D. This data has more data at the low and high extremes and less in the middle, so the "middle 50%" will be more spread out. (It will take a wider "net" to catch the middle 50%)

5. Which data set appears to have a larger mean absolute deviation (or do they appear to be the same)? Why?

Data Set D. Again, this data has more "weirdos" at the high and low extremes, which have higher deviations than the data in the middle, thus increasing the MAD.